THE EFFECT OF TWO CONTRASTING FORMS OF TESTING UPON LEARNING

A DISSERTATION

SUBMITTED TO THE BOARD OF UNIVERSITY STUDIES OF THE JOHNS HOPKINS UNIVERSITY IN CONFORMITY WITH THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF EDUCATION

BY

SISTER FELICITA GABLE

(Reprinted from "The Johns Hopkins University Studies in Education," Number 25)

BALTIMORE
THE JOHNS HOPKINS PRESS
1936

THE EFFECT OF TWO CONTRASTING FORMS OF TESTING UPON LEARNING

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547018_0005

THE EFFECT OF TWO CONTRASTING FORMS OF TESTING UPON LEARNING

A DISSERTATION

SUBMITTED TO THE BOARD OF UNIVERSITY STUDIES OF THE JOHNS HOPKINS UNIVERSITY IN CONFORMITY WITH THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF EDUCATION

BY

SISTER FELICITA GABLE

(Reprinted from "The Johns Hopkins University Studies in Education," Number 25)

BALTIMORE

THE JOHNS HOPKINS PRESS

1936

Copyright 1936, The Johns Hopkins Press

Printed in the United States of America
by J. H. Furst Company, Baltimore, Maryland

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Dr. John M. Stephens, to Dr. Florence E. Bamberger, to Dr. E. Earle Franklin, and to Dr. Arthur Lichtenstein of the Johns Hopkins University for the valuable suggestions given in the preparation of this dissertation.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	THE PROBLEM	1
	Introduction	
	Review of Research in the Field	
	The Problem and its Setting	
II.	PROCEDURE	7
	Time of Experimentation	
	Selection of Subjects	
	Equating of Groups	
	Technique	
	Identical Factors	
	Precautions against Spread of Information	
	Points of Difference between the Present Investigation and the Research Reviewed in Chapter I	
III.	RESULTS	18
	Major Findings	
	Minor Findings	
	Suggested Problems of Research	
IV.	CONCLUSION	28
V.	SUMMARY	29
BIBLIOGRAPHY		31
INDEX		33
VITA		35

LIST OF TABLES

TABLE		PAGE
I.	Frequency Distributions for Mental Age—One Day Absence—First and Second Experiments	8
II.	Frequency Distributions for Initial Scores—One Day Absence—First and Second Experiments	9
III.	Frequency Distributions for Socio-Economic Scores—One Day Absence—First and Second Experiments	10
IV.	Frequency Distributions for Mental Age—No Absence—First and Second Experiments	11
V.	Frequency Distributions for Initial Scores—No Absence—First and Second Experiments	12
VI.	Frequency Distributions for Socio-Economic Scores—No Absence—First and Second Experiments	13
VII.	Frequency Distributions for Final Scores—One Day Absence—First and Second Experiments	18
VIII.	Comparison of Equating Factors and Gains—One Day Absence—First Experiment	19
IX.	Comparison of Equating Factors and Gains—One Day Absence—Second Experiment	19
X.	Frequency Distributions for Final Scores—No Absence—First and Second Experiments	20
XI.	Comparison of Equating Factors and Gains—No Absence—First Experiment	21
XII.	Comparison of Equating Factors and Gains—No Absence—Second Experiment	21
XIII.	Frequency Distributions for Mental Age—March 16	22
XIV.	Frequency Distributions for Initial Scores—March 16	23
XV.	Frequency Distributions for Socio-Economic Scores—March 16	23
XVI.	Frequency Distributions for Final Scores—March 16	24
XVII.	Comparison of Equating Factors and Gains—March 16.	25
XVIII.	Comparison of Gains—Highest Third—One Day Absence—First Experiment.	25
XIX.	Comparison of Gains—Middle Third—One Day Absence—First Experiment	26
XX.	Comparison of Gains—Lowest Third—One Day Absence—First Experiment	26
XXI.	Differences in Gains—First and Second Experiments	27

CHAPTER I

The Problem

INTRODUCTION

Procedures concerning the administration of tests and the frequency of testing have attracted the attention of educators and have provided the subject of much controversy and discussion. Modern philosophy impresses one that education is life and growth, and not a mere matter of accumulating facts. However, educators seem to be agreed that pupils tend to accomplish more when confronted with the realization that a day of reckoning is at hand when they are expected to give an account of their knowledge. Such a situation contains dynamic or motivating properties which aid the crucial aim in teaching—motivation. "Starch considers knowledge of success or failure one of the potent factors in improvement in school subjects."[1] Knowledge of success or failure can be conveyed in no more systematic manner than through testing, which, therefore, is an important field for investigation.

REVIEW OF RESEARCH IN THE FIELD

In reviewing research in the field the experimenter observed that tests, frequent and infrequent, announced and unannounced, have been problems of investigators.

In 1929, Deputy, Teachers College, Columbia University, used a class in freshman philosophy. At the first meeting of the class his plan was exposed to the students. The subjects were matched on intelligence. Of three sections of 35 students each: Section I was given a ten-minute check at each bi-weekly meeting of the class; Section II was the control for the first half-semester and received no checks; Section III was given a twenty-minute check once a week on the previous lesson. In order to equalize the opening for spread of information, it was arranged that every Friday half of

[1] Deputy, E. C., Knowledge of Success as Motivating Influence in College Work, J. Ed. Research, 20 (1929), 329.

the written work of Section III was the material that Section I had had on Thursday and the other half was material that Section I would have the following week. "The first ten minutes of each meeting in (Section II) were devoted to reviewing, orally, the preceding lesson by discussing the points contained in the written work of Section I and III. . . . In case a student was absent, he was required to take the written exercises before his excuse card was signed.[2] . . . At the time of the mid-semester examination the experimental sections had had six weeks of the written and recitation work as described above." [3] During the latter half of the semester Section II became the experimental while the other groups received no checks. In both experiments the control group or groups failed to achieve the average of the experimental sections.

In 1930, Pease,[4] College of the Pacific, included 204 students in high school and college classes in an investigation. He divided them into two groups on the basis of intelligence. The first group took a 100-item objective test without warning. The second was dismissed for cramming in anticipation of receiving the same 100-item test at the next meeting of the class. (This set was told to cram for at least one hour during the interim, and on the day that they were given the test the total cramming time was recorded.) Six weeks later the identical test was administered to both groups. The results of the first testing showed an experimental gain of 11.1 points over the control; the experimental gain in the second test was 6.3.

Pease conducted successive experiments of this type with various subjects. In one class of 94, with a mean intelligence of 139.1, the experimental gained over the control on the initial test 8.2 ± 2.76, and 6 weeks later 4.8 ± 2.25. While in another class of 82, with a mean intelligence of 163.4, the experimental gain on the initial test was 17.07 ± 3.32 and 12 weeks later was 2.7 ± 3.17.

[2] Deputy, E. C., Knowledge of Success as Motivating Influence in College Work, J. Ed. Research, 20 (1929), 329.
[3] *Ibid.*, 330.
[4] Pease, G. R., Should Teachers Give Warning of Tests and Examinations? J. Ed. Psychol., 21 (1930), 273-77.

In 1931, Turney,[5] University of Kansas, selected as an experimental group a section of 40 students in educational psychology of slightly lower intelligence than the control group of 28. The former had a mean pretest score of 85.2 versus 108.1 for the control. The experimental group were told that they would have a short quiz about every week. No information concerning testing was given to the control group. (They received no quizzes.) Twelve were given throughout the semester. "The scores were determined promptly after each quiz and reported to the class at the next class meeting."[6] Despite the handicap on initial scores, the experimental group scored about equal to the control on the final examination.

In the spring semester of 1934, Noel Keys,[7] University of California, selected from his class in educational psychology two groups of 143 students each. He matched them on sex and pretest scores. The fifteen weeks of experimentation were divided into three four-week periods. For the first period the control were given a lump assignment together with the date of the mid-term test; while the experimental set were assigned weekly readings by topics and chapters, and the dates for the tests each week were specified. For the second period the same weekly assignments were given to both groups, but the difference in testing procedure continued as in the first period. For the third period, weekly assignments were given only to the experimental group, but the groups received one monthly test. Both groups received the same test material and the same amount of test material. "The control section, however, met after the experimental and probably enjoyed a certain advantage on that account. To offset this as far as practicable, one-fourth of the lectures and one half of the total tests and examinations were scheduled to be given first to the control. A further possible advantage to the control group lay in the fact that 14% of

[5] Turney, A. H., The Effect of Frequent Short Objective Tests upon the Achievement of College Students in Educational Psychology, Sch. and Soc., 33 (1931), 760-62.

[6] *Ibid.*, 761.

[7] Keys, Noel, The Influence on Learning and Retention of Weekly as Opposed to Monthly Tests, J. Ed. Psychol., 25 (1934), 427-36.

its members were graduate students, as against 6% of the experimental." [8] The total mean gain of the weekly-tested group over the monthly-tested group on the periodic test material was 26.5 ± 4.1. The difference in gains in scores on an unannounced final examination after a lapse of 5 to 14 weeks was 4.2 ± 1.4 in favor of the weekly-tested group. However, on the regular end-term examination where all had equal opportunity to cram, scores were the same for both groups.

Noel Keys reports an investigation by C. H. Smeltzer. Smeltzer used two educational psychology groups numbering 26 and 28, respectively. " The experimental section was given weekly tests each Thursday and the papers returned and ' gone over ' in Friday class." [9] Adjustments were made for differences in pretest scores. Both groups were given a final examination on which the median score for the experimental group was 230.6, median score for the control group was 219.3.

In 1934, Weissmann,[10] College of the City of New York, studied the effect of giving one group daily tests for eight to nine weeks, while a second group received no tests. The number of subjects for the groups, respectively, was 181 and 180, students in high-school physics. The author says, " Since no intelligence tests were administered it is difficult to compare the two groups as far as learning ability is concerned. The results are therefore limited by the assumption that the chronological ages, previous averages in mathematics, previous averages in physics, and the scores obtained on the initial test are sufficient as basis for equating the groups." [11] Samples of the daily tests are not shown in the manuscript copy. Although many of the items on initial and final tests are identical, " The initial test consisted of 30 true-false statements, and 30 completions. . . . The final test consisted of three parts. Part I consisted of 30 true-false statements, Part II consisted of

[8] *Ibid.*, 429.
[9] *Ibid.*, 428.
[10] Weissman, S. A., The Effect of Frequent Testings on Achievement in High School Physics, College of the City of N. Y., Master's (1934), 41 p. ms.
[11] *Ibid.*, 14.

30 completion items and Part III consisted of 10 problems to be solved." [12] With such change in the measuring "instrument" Weissman records a gain of 6.96 ± 0.89 for the daily-test group over the no-test group.

So much for the matter of frequency, now a few gleanings concerning procedures will be presented. In 1923, Jones,[13] Columbia University, found that written examinations given at the close of a presentation increased retention. In summary: "Five series of experiments were conducted on paired associates, lecture samples, and hour lectures, to investigate the effect of examination on classroom learning, as measured after intervals from 3 days to 8 weeks. In all experiments, the retest scores were reliably higher than the control scores, the retest average ranging from 33% to approximately 100% above the control average." [14]

In 1935, Corey,[15] University of Nebraska, compared correlations of Army Alpha scores of 104 students with test results obtained from surprise quizzes, on the one hand, and on the other hand with results obtained from a final examination announced long in advance. His assumption is that achievement is motivated much more adequately in a final examination than in a surprise quiz.

And so results of studies, from 1923 to 1935, appear to be in favor of frequent and announced tests. The question that presented itself to the present experimenter was: But just how frequently should tests be given? Is there a certain desirable "frequency"? And it was near the close of the summer of 1935 that she was acquainted with the fact that the teachers of biology in the Seton High School, Baltimore, had introduced in November, 1934, regular daily check as a supplement to unit testing. Seton enrolls 1300 girls from all

[12] *Ibid.*, 7.

[13] Jones, J. E., Experimental Studies of College Teaching, Sections IV and V, Archives of Psychol., No. 68 (1923), 36-54.

[14] *Ibid.*, 51.

[15] Corey, S. M., The Effect of Motivation upon the Relationship between Achievement and Intelligence, Sch. and Soc., 41 (1935), 256-57.

sections of the city. The number enrolled in the biology department is 365. Here was the setting for the problem.

To what extent would a consistent check from day to day in itself suffice as a spur to achievement? Again, could pupils of the ninth grade be so highly motivated by an initial announcement as to keep themselves prepared for a series of irregular surprise tests over a prolonged period of teaching? In contrast, what motivation is supplied by immediate announcement of more or less comprehensive tests at irregular intervals of not less than five, not more than fifteen days, with time given for review?

And so the problem simmered and was thus finally stated: Does a System of "Sprung" Testing Effect Greater Achievement in Ninth-Grade Biology than a System of "Anticipated Daily Check"[16] Testing?

This daily check group received consistently a check of from ten to twenty questions covering the work of the previous day or days. These tests, approximating three a week, were given at the beginning of each lecture period. On the other hand, the sprung group received no check tests but were informed at the beginning of the experimental period that "until further notice you will be given tests at various uneven intervals without warning, on the matter covered to date of test." Dates of sprung tests were so timed that they sometimes overlapped the conclusion of a unit. The control group were given no information at the outset, and had no check tests. But on the day that any test was administered to the sprung group, the control group were told that they would receive a test on a specified date. Consequently, they were given the same test as the sprung group after a lapse of not more than one school day. The same material was used to make up both the daily check and sprung tests.

[16] For the ease of the reader the "Anticipated Daily Check" group will be termed "Daily Check" group throughout this paper.

CHAPTER II

Procedure

TIME OF EXPERIMENTATION

The period of experimentation was from October 29, 1935, to February 26, 1936. The first experiment extended from October 29 to December 19. A second experiment began January 6 using the same subjects but reversing groups for the daily check and sprung procedures.

SELECTION OF SUBJECTS

Each experiment was carried on by two teachers of the biology department, each teaching three sections totaling 274 students. The number of subjects for the daily check group of the first seven weeks was 99; for the sprung group 100; and for the control group 75. Each teacher had a section running on each testing plan. A pupil was not excluded from the experiment for one day's absence * provided no test was missed; all other absence caused elimination.

EQUATING OF GROUPS

The pupils were equated on intelligence, results of a pretest in biology and socio-economic status. Mental ages in months were determined from raw scores on the Otis Self-Administering Tests of Mental Ability, Higher Examination: Form A. The pretest for each experiment consisted of 100 items of the objective type. The matter of these tests was selected from the Presson Biology Tests and the Ruch-Cossman Biology Tests, publications of the World Book Company. This material was "reproduced by written permission of the publishers." (Without warning to any group, these tests were reissued to all groups as final tests.)

For measuring the third factor, the Sims Score Card for Socio-Economic Status was used. The publishers state: "The correlation between two groups of siblings for the

* To offset absence in the second experiment, due to unusually severe weather conditions, the experimenter deemed it advisable to allow one day to any pupil throughout the course of each study.

questions used in the Score Card was found to be .95 ± 0.01." Tables I, II, and III show the frequency distributions, means, standard deviations, and probable errors for the equating factors, respectively. The differences between mean scores on mental age, pretest, and socio-economic status are all statistically insignificant. The greatest difference is 1.0 ± 1.2 for socio-economic averages, obtained by subtracting the daily check from the sprung, and also by subtracting the control from the sprung. However, this difference is very much less than four times its probable error. Nomographs * were used to calculate probable errors.

TABLE I

Frequency Distributions for Mental Age
Allowing Absence of One Day—No Test Missed

	FIRST EXPERIMENT			SECOND EXPERIMENT		
Mental Age (in months)	S Group	ADC Group	C Group	S Group	ADC Group	C Group
204.5–209.5	1		1		1	
199.5–204.5	1	2	1	2		1
194.5–199.5	2	3	1	3	1	
189.5–194.5	5	5	1	2	4	
184.5–189.5	2	2	3	1	1	3
179.5–184.5	7	3	3	1	5	4
174.5–179.5	2	4	8	2	3	6
169.5–174.5	9	8	8	5	6	5
164.5–169.5	8	10	5	3	6	4
159.5–164.5	5	5	7	2	2	4
154.5–159.5	8	3	1	1	2	2
149.5–154.5	6	5	1	2	2	
144.5–149.5	2	4	5	4	1	2
139.5–144.5	5	1	5	2	3	3
134.5–139.5	4	2	1	2	3	1
129.5–134.5	3	4	2	2	2	1
124.5–129.5						
119.5–124.5			1		1	1
114.5–119.5		1				
109.5–114.5		1				
Total	70	63	54	34	43	37
Mean	165.4	165.6	165.6	166.4	166.2	166.3
SD	18.5	20.0	18.0	20.5	19.5	17.0
PE	1.5	1.5	1.6	2.4	2.0	1.8

N. B. Throughout the tables of this Dissertation: S = Sprung
ADC = Daily Check
C = Control

* Dunlap & Kurtz, Handbook of Statistical Nomographs, Tables, and Formulas, World Book Co., 1932.

TABLE II

FREQUENCY DISTRIBUTIONS FOR INITIAL SCORES ALLOWING ABSENCE OF ONE DAY—NO TEST MISSED

FIRST EXPERIMENT				SECOND EXPERIMENT			
Initial	S Group	ADC Group	C Group	Initial	S Group	ADC Group	C Group
36.5–38.5		1					
34.5–36.5							
32.5–34.5		2		33.5–35.5	1		
30.5–32.5	2	1		31.5–33.5	1		
28.5–30.5		1	1	29.5–31.5		1	
26.5–28.5	1	1	1	27.5–29.5	2		1
24.5–26.5	4	3	3	25.5–27.5	2		1
22.5–24.5	7	3	10	23.5–25.5		3	4
20.5–22.5	3	8	3	21.5–23.5	1	3	4
18.5–20.5	18	7	13	19.5–21.5	3	5	1
16.5–18.5	12	10	5	17.5–19.5	3	6	3
14.5–16.5	9	10	5	15.5–17.5	3	7	6
12.5–14.5	10	9	7	13.5–15.5	1	4	4
10.5–12.5	2	4	2	11.5–13.5	5	3	3
8.5–10.5	1	2	2	9.5–11.5	6	6	3
6.5– 8.5	1	1	1	7.5– 9.5	5	2	5
4.5– 6.5				5.5– 7.5	1	2	2
2.5– 4.5			1	3.5– 5.5			
				1.5– 3.5		1	
Total	70	63	54		34	43	37
Mean	18.5	18.6	18.5		16.3	16.1	16.2
SD	4.6	6.0	5.2		7.4	5.6	6.0
PE	0.4	0.5	0.5		0.8	0.6	0.7

TABLE III

Frequency Distributions for Socio-Economic Scores Allowing Absence of One Day—No Test Missed

FIRST EXPERIMENT				SECOND EXPERIMENT			
Socio-Economic	S Group	ADC Group	C Group	Socio-Economic	S Group	ADC Group	C Group
59.5–64.5	3			58.5–62.5	1	2	
54.5–59.5	1	2	1	54.5–58.5			1
49.5–54.5	5	6	3	50.5–54.5	3	1	
44.5–49.5	3	2	4	46.5–50.5	2	3	1
39.5–44.5	7	8	6	42.5–46.5		1	3
34.5–39.5	8	7	14	38.5–42.5	4	5	6
29.5–34.5	22	12	6	34.5–38.5	1	7	12
24.5–29.5	7	13	9	30.5–34.5	6	7	2
19.5–24.5	13	10	7	26.5–30.5	6	8	5
14.5–19.5		3	2	22.5–26.5	8	5	4
9.5– 9.5	1		2	18.5–22.5	2	3	1
				14.5–18.5	1		2
				10.5–14.5		1	
Total	70	63	54		34	43	37
Mean	34.6	33.6	33.6		33.3	34.2	34.5
SD	10.5	10.0	10.0		10.4	10.0	8.4
PE	0.8	0.8	0.9		1.2	1.0	0.9

As a check on the allowance for one day's absence, groups were also equated by complete elimination for any days missed. The results show insignificant differences for all equating factors as shown in Tables numbers IV, V, and VI; and differences in gains by complete elimination correspond with differences in gains calculated when absence of one day is allowed, as given in Major Findings, Chapter III.

TABLE IV

FREQUENCY DISTRIBUTIONS FOR MENTAL AGE
COMPLETE ELIMINATION FOR ABSENCE

	FIRST EXPERIMENT			SECOND EXPERIMENT		
Mental Age (in months)	S Group	ADC Group	C Group	S Group	ADC Group	C Group
204.5–209.5	1		1			
199.5–204.5	1	1	1	1		1
194.5–199.5	2	3	1	3	1	
189.5–194.5	4	3	2	1	3	
184.5–189.5	2	2	3	1		2
179.5–184.5	5	4	4	1	3	2
174.5–179.5	2	3	7	1	3	4
169.5–174.5	8	11	7	3	5	4
164.5–169.5	8	9	5	5	5	1
159.5–164.5	3	5	7		3	3
154.5–159.5	7	2	2		2	1
149.5–154.5	4	4	1	2	2	
144.5–149.5	1	5	4	5	1	2
139.5–144.5	5	1	5		2	3
134.5–139.5	4	1	1	2	3	1
129.5–134.5	3	3	2	2	2	1
124.5–129.5						
119.5–124.5			1		1	1
114.5–119.5		1				
109.5–114.5		1		1		
Total	60	59	54	28	36	26
Mean	165.2	165.6	166.2	163.2	163.2	163.7
SD	19.0	19.0	18.5	22.5	18.5	19.0
PE	1.5	1.5	1.7	2.8	2.1	2.5

TABLE V

Frequency Distributions for Initial Scores Complete Elimination for Absence

First Experiment				Second Experiment			
Initial	S Group	ADC Group	C Group	Initial	S Group	ADC Group	C Group
36.5–38.5		1					
34.5–36.5							
32.5–34.5		2		33.5–35.5	1		
30.5–32.5	1	1		31.5–33.5			
28.5–30.5			2	29.5–31.5			
26.5–28.5	1	1	1	27.5–29.5	2		
24.5–26.5	3	2	3	25.5–27.5	2		
22.5–24.5	7	3	10	23.5–25.5		1	2
20.5–22.5	2	7	3	21.5–23.5	1	1	2
18.5–20.5	17	7	13	19.5–21.5	1	3	1
16.5–18.5	12	11	5	17.5–19.5	1	6	2
14.5–16.5	5	11	4	15.5–17.5	2	7	3
12.5–14.5	7	6	7	13.5–15.5		4	5
10.5–12.5	3	4	2	11.5–13.5	3	4	3
8.5–10.5		2	2	9.5–11.5	5	6	3
6.5– 8.5	1	1	1	7.5– 9.5	7	1	3
4.5– 6.5	1			5.5– 7.5	3	2	2
2.5– 4.5			1	3.5– 5.5		1	
Total	60	59	54		28	36	26
Mean	18.4	18.4	18.8		14.6	14.8	14.6
SD	4.6	5.8	5.5		7.6	4.6	5.2
PE	0.4	0.5	0.5		1.0	0.6	0.7

TABLE VI

FREQUENCY DISTRIBUTIONS FOR SOCIO-ECONOMIC SCORES
COMPLETE ELIMINATION FOR ABSENCE

FIRST EXPERIMENT				SECOND EXPERIMENT			
Socio-Economic	S Group	ADC Group	C Group	Socio-Economic	S Group	ADC Group	C Group
				62.5–66.5		1	
59.5–64.5	3			58.5–62.5	1		
54.5–59.5		2	1	54.5–58.5	1		1
49.5–54.5	4	5	3	50.5–54.5	2		
44.5–49.5	4	2	4	46.5–50.5	1	3	1
39.5–44.5	6	8	8	42.5–46.5		1	2
34.5–39.5	7	6	14	38.5–42.5	4	4	4
29.5–34.5	15	9	6	34.5–38.5		5	8
24.5–29.5	8	15	8	30.5–34.5	4	6	1
19.5–24.5	12	11	6	26.5–30.5	4	8	4
14.5–19.5		1	2	22.5–26.5	8	5	3
9.5–14.5	1		2	18.5–22.5	2	2	1
				14.5–18.5	1		1
				10.5–14.5		1	
Total	60	59	54		28	36	26
Mean	34.4	33.5	34.3		33.2	33.3	34.8
SD	11.0	10.0	10.0		11.2	9.6	8.8
PE	1.0	0.9	0.9		1.4	1.0	1.2

TECHNIQUE

Every effort was made to maintain the natural situation. It was desired that the students should be unaware of the fact that the various sections were operating on different testing programs. Therefore, with the exception of technical announcements to the sprung group at the opening of the experiments, no information was given to any section or individual student throughout the fourteen weeks. No previous announcement of laboratory periods was made. The student patrol received an immediate order to direct the on-

coming section to the laboratory at the beginning of the regular class change. (Because of this fact, the daily check group were kept in daily anticipation of a test.)

IDENTICAL FACTORS

The same material, namely, teacher-made tests of the preceding year, was used for all daily check and sprung tests. These tests were objective, since the matter of biology is readily adaptable to this type. Of course, the shorter tests measured with the subject-matter taught, and therefore comprised but one or two topics. On the other hand, it was deemed desirable to intermingle topics, and to make the matter of the sprung tests a combined whole. The medium was a mimeographed test on all occasions. All groups received the same number of questions. Testing time was approximately equal in so far as was practical for the maintenance of smooth and non-artificial functioning.

The pupils corrected their daily check, sprung and control tests by a system of exchange. This procedure was used for two reasons: 1. In order that pupils would have knowledge of their success or failure; and 2. That wrong impressions might be immediately corrected. As a matter of fact, Sproule,[1] University of Southern California, states, as a result of his investigation, that even for the true-false examination, the negative effects are overcome by having the children correct their own papers.

The time consumed in the daily check testing was twelve to fifteen minutes. Sprung tests of 50 items were dispatched within 50 minutes. For a 100-item sprung test the correction was carried over to approximately twenty-five minutes of the following day.

The greatest vigilance was exercised to keep instruction identical. Lessons were conducted by the lecture demonstration, lecture recitation, individual-pupil-laboratory, and group laboratory plans. The same lesson plans were used throughout all sections. The same textbook was used, the

[1] Sproule, C. E., Suggestive Effects of the True-False Tests, J. Ed. Psychol., 25 (1934), 284.

assignments were identical, as were the questions dictated as study guides—these questions usually numbered five. Laboratory equipment and procedures were the same for all groups.

The taking and correcting of the daily checks served as a review for these sections. The sprung and control groups, at the beginning of each lecture period, reviewed orally the preceding lesson or lessons by discussing the points covered in the check test of the day. On the day that any sprung test was given the daily check group spent the test-procedure time in reviewing orally the matter covered in the sprung test of the date; these students received no previous announcement of this review. On this day the control received a new lesson, this lesson was given to the sprung and daily check sections on the day of control testing. On the final day of teacher-made testing, for each experiment, both sprung and control received the tests simultaneously—the control group having been warned on the preceding school day. This exception was made in order to " round-up " material. It was considered that the sprung group, not having received their other " surprise " tests on the same day as the control group, would be taken unawares.

PRECAUTIONS AGAINST SPREAD OF INFORMATION

No pupils retained any copies of tests administered. The daily check and sprung tests were written in six different types, namely, Completion, True-False, Matching, Picture Test, Multiple Choice, T and N (true or not true and correct). In making up the sprung tests, care was taken that the numerical order of the daily check tests was disrupted for the sprung. For the most part, the statement of the item was changed, frequently it was recast into a new type of objective statement. The sprung test, therefore, was a remold of the matter of the daily check tests.

The prompt, systematic, economical method of administering, correcting, and collecting tests allowed for little rote memorization of test wording. The system of exchange between pupils varied. As an illustration of one method, papers

were exchanged between one and three, two and four, five and seven, six and eight, whether it be between rows or desks alike. The tests were immediately corrected in unison, whereupon the corrector reversed the paper and signed her name under the words "corrected by"—papers were collected. Only material necessary for the testing procedure was on the desk during the testing.

No student of biology was given any mid-year examination in any subject, so that there was no interference with experimental functioning. Recapitulation was the same throughout the groups. Any point discovered as needing reteaching as a result of daily check tests or tests received general stress in all sections.

The experimenter visited classes Tuesday of each week to observe sections of the various groups taught by the same teacher, and held conferences with the teachers Monday or Friday as necessary. She corrected and scored all the initial and final tests (including those of March 16), and computed and checked all the calculations. It was thought that this would lend to accuracy and equality in the treatment of the three groups.

POINTS OF DIFFERENCE BETWEEN THE PRESENT INVESTIGATION AND THE RESEARCH REVIEWED IN CHAPTER I

1. Three procedures of frequent testing were handled:

A. Brief daily check.

B. A longer test at various uneven intervals, without warning, interval not exceeding two weeks.

C. A longer test at various uneven intervals, announced, interval not exceeding two weeks.

2. These groups were equated on the combined factors of:

A. Intelligence, which is generally admitted and confirmed by practice to be a necessity for scientific method when dealing with subject-matter of this type.

B. Pretest scores, which procedure matched the groups on their initial knowledge of biology thus handicapping none. Moreover, it simultaneously controlled to some extent opportunity for acquiring added knowledge, as they had had such opportunity to acquire such knowledge as they possessed at the outset.

C. Socio-Economic status, by which it was hoped to control home opportunities and facilities which lend aid to the learning process.

3. No other study gave such a bombardment of tests.

4. The children corrected their own papers.

5. The students knew immediately of their success or failure.

6. The experiment was limited to ninth-grade pupils.

CHAPTER III

Results

MAJOR FINDINGS

The findings of the experiments were based on differences in mean gains of the various groups calculated by subtracting average initial scores from average final scores. (The scores on the daily check, sprung, and control tests were not taken as part of the experimental data.)

The major differences for this study are shown in Tables VIII and IX, and the frequency distributions for final scores in Table VII. The control group gained 8.3 ± 1.4 over the daily check group for the first experiment, and 10.6 ± 1.6 for the second experiment; both gains are statistically significant. Control gained over the sprung only 1.5 ± 1.4 for the first, but made a significant gain of 7.5 ± 1.6 for the second. On the other hand, the sprung shows a significant gain of 6.8 ± 1.2 over the daily check in the first experiment but an insignificant gain of 3.1 ± 1.6 in the second experiment.

TABLE VII

Frequency Distributions for Final Scores Allowing Absence of One Day—No Test Missed

	FIRST EXPERIMENT			SECOND EXPERIMENT		
Final	S Group	ADC Group	C Group	S Group	ACD Group	C Group
94.5–99.5		1	1			
89.5–94.5	2	4	5	1		1
84.5–89.5	3	3	6			4
79.5–84.5	16	8	9	4	2	4
74.5–79.5	18	3	6	3	5	5
69.5–74.5	14	6	11	6	5	5
64.5–69.5	4	8	5	2	6	6
59.5–64.5	4	10	6	3	4	6
54.5–59.5	4	5		5	7	4
49.5–54.5		7	3	2	4	
44.5–49.5	2	4	1	5	4	1
39.5–44.5	1	1		1	1	1
34.5–39.5	1	1	1	2	4	
29.5–34.5	1	1				
24.5–29.5						
19.5–24.5		1			1	
Total	70	63	54	34	43	37
Mean	72.8	66.2	74.4	63.0	59.7	70.4
SD	11.5	15.5	12.0	14.0	13.5	11.5
PE	0.9	1.1	1.1	1.6	1.4	1.3

TABLE VIII

COMPARISON OF EQUATING FACTORS AND GAINS (FIRST EXPERIMENT) ALLOWING ABSENCE OF ONE DAY—NO TEST MISSED

		S Group		ADC Group		C Group		S − ADC		S − C		C − ADC	
		Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
MA	Mean	165.4	1.5	165.6	1.5	165.6	1.6	−0.2	2.1	−0.2	2.2	0.0	2.2
	SD	18.5		20.0		18.0							
CA	Mean	175.7	0.7	173.6	0.7	175.1	0.8	2.1	1.0	0.6	1.3	1.5	1.0
	SD	9.0		7.8		8.4							
SES	Mean	34.6	0.8	33.6	0.8	33.6	0.9	1.0	1.2	1.0	1.2	0.0	1.2
	SD	10.5		10.0		10.0							
F	Mean	72.8	0.9	66.2	1.1	74.4	1.1						
	SD	11.5		15.5		12.0							
I	Mean	18.5	0.4	18.6	0.5	18.5	0.5	−0.1	0.6	0.0	0.6	−0.1	0.6
	SD	4.6		6.0		5.2							
Gain		54.3	0.8	47.6	0.9	55.9	1.1	6.8	1.2	−1.5	1.4	8.3	1.4
		(r = .43)		(r = .35)		(r = .11)							

N. B. Throughout the Tables of Comparison of this Dissertation: MA = Mental Age; CA = Chronological Age; SES = Socio-Economic status; F = Final; and I = Initial.

TABLE IX

COMPARISON OF EQUATING FACTORS AND GAINS (SECOND EXPERIMENT) ALLOWING ABSENCE OF ONE DAY—NO TEST MISSED

		S Group		ADC Group		C Group		S − ADC		S − C		C − ADC	
		Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
MA	Mean	166.4	2.4	166.2	2.0	166.3	1.8	0.2	3.9	0.1	3.0	0.1	2.7
	SD	20.5		19.5		17.0							
SES	Mean	33.3	1.2	34.2	1.0	34.5	0.9	−0.9	1.6	−1.2	1.5	0.3	1.4
	SD	10.4		10.0		8.4							
F	Mean	63.0	1.6	59.7	1.4	70.4	1.3						
	SD	14.0		13.5		11.5							
I	Mean	16.3	0.8	16.1	0.6	16.2	0.7	0.2	1.0	0.1	1.9	0.1	0.8
	SD	7.4		5.6		6.0							
Gain		46.7	1.2	43.6	1.1	54.2	1.2	3.1	1.6	−7.5	1.6	10.6	1.6
		(r = .66)		(r = .61)		(r = .31)							

For the purpose of comparison, differences calculated for those equated groups which excluded for any absence are shown in Tables XI and XII, and the frequency distributions for final scores in Table X. In both the first and second experiments the control minus the daily check gives significant gains of 7.1 ± 1.7, and 8.5 ± 1.9, respectively. The gains of the control over the sprung are 2.7 ± 1.3 for the first, and 4.0 ± 2.0 for the second, both insignificant. On the other hand, sprung minus daily check gives insignificant but considerable differences in favor of the sprung, 4.4 ± 1.4 for the first experiment, and 4.6 ± 1.7 for the second.

TABLE X

Frequency Distributions for Final Scores
Complete Elimination for Absence

	FIRST EXPERIMENT			SECOND EXPERIMENT		
Final	S Group	ADC Group	C Group	S Group	ADC Group	C Group
94.5–99.5		1	1			
89.5–94.5	1	3	5	1		1
84.5–89.5	3	4	7			2
79.5–84.5	12	6	10	3	2	1
74.5–79.5	15	5	7	3	2	3
69.5–74.5	13	5	8	4	3	3
64.5–69.5	3	9	5	2	6	3
59.5–64.5	5	10	6	3	4	6
54.5–59.5	3	5		3	8	4
49.5–54.4		6	3	3	4	1
44.5–49.5	2	2	1	3	1	1
39.5–44.5	1	1		1	1	1
34.5–39.5	1	1	1	2	4	
29.5–34.5	1	1				
24.5–29.5						
19.5–24.5					1	
Total	60	59	54	28	36	26
Mean	71.9	67.5	75.0	62.8	58.5	66.8
SD	12.0	14.0	12.5	14.0	13.5	12.0
PE	1.0	1.2	1.2	1.8	1.5	1.6

TABLE XI

Comparison of Equating Factors and Gains
(First Experiment)
Complete Elimination for Absence

		S Group		ADC Group		C Group		S − ADC		S − C		C − ADC	
		Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
MA	Mean	165.2	1.5	165.6	1.5	166.2	1.7	−0.4	2.2	−1.0	2.3	0.6	2.3
	SD	19.0		19.0		18.5							
SES	Mean	34.4	1.0	33.5	0.9	34.3	0.9	0.9	1.2	0.1	1.3	0.8	1.2
	SD	11.0		10.0		10.0							
F	Mean	71.9	1.5	67.5	1.2	75.0	1.2						
	SD	12.0		14.0		12.5							
I	Mean	18.4	0.4	18.4	0.5	18.8	0.5	0.0	0.6	−0.4	0.6	0.4	0.7
	SD	4.6		5.8		5.5							
Gain		53.5	0.6	49.1	1.3	56.2	1.2	4.4	1.4	−2.7	1.3	7.1	1.7
		(r = .55)		(r = .24)		(r = .16)							

TABLE XII

Comparison of Equating Factors and Gains (Second Experiment)
Complete Elimination for Absence

		S Group		ADC Group		C Group		S − ADC		S − C		C − ADC	
		Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
MA	Mean	163.2	2.8	163.2	2.1	163.7	2.5	0.0	3.5	−0.5	3.8	0.5	3.2
	SD	22.5		18.5		19.0							
SES	Mean	33.2	1.4	33.3	1.0	34.8	1.2	−0.1	1.8	−1.6	1.8	1.5	1.6
	SD	11.2		9.6		8.8							
F	Mean	62.8	1.8	58.5	1.5	66.8	1.6						
	SD	14.0		13.5		12.0							
I	Mean	14.6	1.0	14.8	0.6	14.6	0.6	−0.2	1.1	0.0	1.2	−0.2	0.8
	SD	7.6		4.6		5.2							
Gain		48.2	1.3	43.7	1.2	52.2	1.5	4.6	1.7	−4.0	2.0	8.5	1.9
		(r = .70)		(r = .58)		(r = .28)							

MINOR FINDINGS

In the course of the experimental procedure the question had arisen: Did the control and sprung groups have an advantage over the daily check group in that they received the bulk of the material of the regular class procedure tests closer to the finals, an advantage of recency? For the purpose of supplying the best possible answer to this incidental problem the final test of December 19 was repeated on March 16. Due to absence it was necessary, for the retesting, to re-equate the daily check and control groups. The sprung group, however, retained its original number of 70 subjects as equated for the first experiment. The frequency distributions for the equating factors are presented in Tables XIII, XIV, and XV.

TABLE XIII

FREQUENCY DISTRIBUTIONS FOR MENTAL AGE (MARCH 16) ALLOW ABSENCE OF ONE DAY DURING THE FIRST EXPERIMENT

Mental Age (in months)	S Group	ADC Group	C Group
204.5–209.5	1		1
199.5–204.5	1	2	1
194.5–199.5	2	3	1
189.5–194.5	5	3	1
184.5–189.5	2	2	3
179.5–184.5	7	3	2
174.5–179.5	2	4	8
169.5–174.5	9	11	7
164.5–169.5	8	8	4
159.5–164.5	5	5	6
154.5–159.5	8	2	1
149.5–154.5	6	5	1
144.5–149.5	2	5	4
139.5–144.5	5	1	5
134.5–139.5	4	2	1
129.5–134.5	3	4	2
124.5–129.5			
119.5–124.5			1
114.5–119.5			
109.5–114.5		1	
Total	70	61	49
Mean	165.4	165.6	165.6
SD	18.5	19.0	18.5
PE	1.5	1.6	1.8

TABLE XIV

FREQUENCY DISTRIBUTIONS FOR INITIAL SCORES (MARCH 16) ALLOW ABSENCE OF ONE DAY DURING THE FIRST EXPERIMENT

Initial	S Group	ADC Group	C Group
36.5–38.5		1	
34.5–36.5			
32.5–34.5		2	
30.5–32.5	2	1	
28.5–30.5			1
26.5–28.5	1	1	1
24.5–26.5	4	3	2
22.5–24.5	7	3	10
20.5–22.5	3	8	3
18.5–20.5	18	6	11
16.5–18.5	12	10	5
14.5–16.5	9	12	4
12.5–14.5	10	8	7
10.5–12.5	2	4	2
8.5–10.5	1	2	2
6.5– 8.5	1		1
Total	70	61	49
Mean	18.5	18.6	18.7
SD	4.6	5.6	4.8
PE	0.4	0.5	0.5

TABLE XV

FREQUENCY DISTRIBUTIONS FOR SOCIO–ECONOMIC SCORES (MARCH 16) ALLOW ABSENCE OF ONE DAY DURING THE FIRST EXPERIMENT

Socio-Economic	S Group	ADC Group	C Group
59.5–64.5	3		
54.5–59.5	1	2	1
49.5–54.5	5	6	3
44.5–49.5	3	2	4
39.5–44.5	7	7	5
34.5–39.5	8	7	13
29.5–34.5	22	11	6
24.5–29.5	7	12	9
19.5–24.5	13	10	5
14.5–19.5		3	2
9.5–14.5	1	1	1
Total	70	61	49
Mean	34.6	33.3	34.3
SD	10.5	10.5	9.5
PE	0.8	0.9	0.9

The repetition of this final test, after a lapse of approximately three months, showed gain differences of 1.0 ± 1.6 in favor of the control over the sprung; 2.9 ± 1.2 in favor of the sprung over the daily check; and 3.9 ± 1.6 in favor of the control over the daily check. As with the findings of other experiments, that of Pease,[1] and that of Keys,[2] summarized in Chapter I of the present write-up, retest differences seem to decrease and to become statistically insignificant. The frequency distributions for final scores are given in Table XVI; the comparison of equating factors and gains in Table XVII.

TABLE XVI

Frequency Distributions for Final Scores (March 16) Allow Absence of One Day During the First Experiment

Final	S Group	ADC Group	C Group
89.5–94.5	2		4
84.5–89.5	3	4	2
79.5–84.5	7	3	3
74.5–79.5	14	8	7
69.5–74.5	9	9	9
64.5–69.5	13	10	9
59.5–64.5	6	4	4
54.5–59.5	3	11	2
49.5–54.5	3	4	4
44.5–49.5	4	3	3
39.5–44.5	3	4	2
34.5–39.5	2		
29.5–34.5	1	1	
Total	70	61	49
Mean	67.6	64.8	68.8
SD	13.5	12.5	13.0
PE	1.1	1.1	1.2

[1] Pease, G. R., Should Teachers Give Warning of Tests and Examinations? J. Ed. Psychol., 21 (1930), 273-77.

[2] Keys, N., The Influence on Learning and Retention of Weekly as Opposed to Monthly Tests, J. Ed. Psychol., 25 (1934), 427-36.

TABLE XVII

COMPARISON OF EQUATING FACTORS AND GAINS (MARCH 16) ALLOW ABSENCE OF ONE DAY DURING FIRST EXPERIMENT

		S Group		ADC Group		C Group		S — ADC		S — C		C — ADC	
		Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
MA	Mean	165.4	1.5	165.6	1.6	165.6	1.8	−0.2	2.2	−0.2	2.3	0.0	2.4
	SD	18.5		19.0		18.5							
SES	Mean	34.6	0.8	33.3	0.9	34.3	0.9	1.3	1.2	0.3	1.2	1.0	1.2
	SD	10.5		10.5		9.5							
F	Mean	67.6	1.1	64.8	1.1	68.8	1.2						
	SD	13.5		12.5		13.0							
I	Mean	18.5	0.4	18.6	0.4	18.7	0.4	−0.1	0.6	−0.2	0.6	0.1	0.6
	SD	4.6		5.6		4.8							
Gain		49.1 (r = .39)	0.9	46.2 (r = .42)	0.9	50.1 (r = .02)	1.3	2.9	1.2	−1.0	1.6	3.9	1.6

Again, by dividing the three groups of the first experiment into thirds on the basis of group equation on intelligence, allowing absence of one day, provided no test was missed, the investigator made an interesting observation. For all three divisions the order prevails, control first, sprung second, and daily check third. For the highest third the differences are very slight. For the middle third differences are worthy of note. For the lowest third the control group appears to outstrip the daily check group by considerable leads. Below are the tables:

TABLE XVIII

COMPARISON OF GAINS HIGHEST THIRD

		FIRST EXPERIMENT						ONE DAY ABSENCE					
		S Group		ADC Group		C Group		S — ADC		S — C		C — ADC	
		Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
F	Mean	79.7	0.8	78.4	1.8	81.1	1.5						
	SD	5.4		12.3		9.6							
I	Mean	20.0	0.6	21.4	1.0	20.9	0.7						
	SD	4.0		7.2		4.4							
Gain		59.7 (r = .08)	0.9	57.0 (r = .41)	0.8	60.2 (r = .56)	1.0	2.7	1.2	−0.5	1.3	3.2	1.3

TABLE XIX

Comparison of Gains
Middle Third

		FIRST EXPERIMENT			ONE DAY ABSENCE		
		S Group	ADC Group	C Group	S — ADC	S — C	C — ADC
		Value PE	Value PE	Value PE	Value PE	Value PE	Value PE
F	Mean	74.2 1.1	66.7 1.6	76.2 1.2			
	SD	8.4	11.1	7.5			
I	Mean	19.2 0.6	17.8 0.7	17.6 0.8			
	SD	4.6	4.8	5.4			
Gain		55.0 1.1 (r = .34)	48.9 1.3 (r = .23)	58.6 1.4 (r = .03)	6.1 1.7	−3.6 1.8	9.7 1.9

TABLE XX

Comparison of Gains
Lowest Third

		FIRST EXPERIMENT			ONE DAY ABSENCE		
		S Group	ADC Group	C Group	S — ADC	S — C	C — ADC
		Value PE	Value PE	Value PE	Value PE	Value PE	Value PE
F	Mean	63.7 1.9	54.3 1.7	66.3 0.7			
	SD	13.5	11.7	4.6			
I	Mean	16.3 0.6	16.6 0.6	16.9 0.7			
	SD	4.0	4.4	4.4			
Gain		47.4 1.8 (r = .36)	37.7 1.6 (r = .35)	49.4 1.0 (r = .05)	9.7 2.4	−2.0 2.0	11.7 1.9

SUGGESTED PROBLEMS OF RESEARCH

The above data would lead one to suspect that research to discover what procedures and methods would best serve for varying intellectual levels might lead to some important findings.

Incidentally, an experiment to determine the number of days of warning that should be given might be worth discovering.

Finally, a study to find out which of two procedures—announced unit tests at uneven intervals with daily oral review versus announced unit tests at uneven intervals with daily written tests—appears to be a fruitful source of experimentation.

As a final review of the results of the present investigation, a comparison of differences in gains might prove interesting.

TABLE XXI

DIFFERENCES IN GAINS

	FIRST EXPERIMENT AND SECOND EXPERIMENT											
	S Group		ADC Group		C Group		S / ADC		S / C		C / ADC	
	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE	Value	PE
Allowing One Day Absence	6.8	1.2	−1.5	1.4	8.3	1.4	3.1	1.6	−7.5	1.6	10.6	1.6
Allowing No Absence	4.4	1.4	−2.7	1.3	7.1	1.7	4.6	1.7	−4.0	2.0	8.5	1.9
Retest March 16	2.9	1.2	−1.0	1.7	3.9	1.7						

CHAPTER IV

Conclusion

In view of the fact that the control group at no time lost its lead, though in the first experiment its gain over the sprung group was very slight, it appears that a series of opportunities for review throughout the learning period are productive of somewhat superior retention.

On the other hand, the writer considers that sprung testing deserves consideration, since in the one instance of the first experiment which is based on the greatest number of cases its gain over the daily check group was significant, namely, 6.8 ± 1.2. The writer thinks that the sprung group must have gained a certain desirable something which her tests did not measure. Perhaps at some future time more adequate measures will be forthcoming to determine if pupils so trained gain in responsibility and self-discipline when upsupported by teacher-reminders of tests, or by daily testing.

As to the daily check system, it appears that this system when compared with a system of giving longer tests, announced and unannounced, at intervals of not less than one week and not more than two weeks, supplies insufficient motivation. Perhaps it is stating it more accurately to say that motivation does not persist.

Again, it appears that allowing absence of one day had no weighty effect on differences in average gains. This seems to be in accord with one of the conclusions of Stephens,[1] Johns Hopkins University. He, in his elaborate thesis, amassed pertinent data gathered by numerous investigators to show that "The relationship between school attendance and achievement is positive but surprisingly slight. (The correlation ranges from 0 to 0.20, approximately.)"

[1] Stephens, J. M., The Influence of the School on the Individual, Ann Arbor, Mich., Edwards Bros. 1933, p. 1.

CHAPTER V

Summary

The present study attempts to determine the effect on pupil achievement of a system of anticipated daily check testing as compared with frequent announced unit tests, and frequent unannounced unit tests. The same volume of test material was used and approximately the same time was consumed in testing. The findings were based on the performance of ninth-grade biology students, numbering ninety-nine daily check, one hundred unannounced, and seventy-five announced. Two experiments were conducted using the same subjects. In the second experiment, the testing procedure was reversed for two of the groups. Groups were equated for mental age, initial knowledge of biology, and socio-economic status.

Results indicate that, under the conditions of this investigation:

1. It is quite certain that announced and unannounced unit tests given at longer intervals effect greater achievement as measured by a final test given without warning, than does a system of daily tests. It is less sure that giving pupils a series of opportunities for review throughout the learning period is productive of greater retention as measured by a final unannounced test.
2. Awareness of the fact that tests will be given unexpectedly suffices to keep students "with candles lighted," since the sprung group attained a relatively high achievement.
3. Unannounced retesting after a lapse of three months shows that the groups who had received unit tests throughout the teaching period maintain their leads, with differences decreasing.
4. For brighter pupils, it makes slight difference which of the testing procedures is used.
5. For the slower pupils, a warning of a test with time to review the whole appears to be very effective, and an unannounced test on the whole slightly less effective.

BIBLIOGRAPHY

1. Books

1. Cole, W. E., The Teaching of Biology, N. Y., D. Appleton-Century Co., 1934.
2. Stephens, J. M., The Influence of the School on the Individual, Ann Arbor, Mich., Edwards Bros., 1933.

2. Periodicals

3. Corey, S. M., "The Effect of Motivation upon the Relationship between Achievement and Intelligence," Sch. and Soc., 41 (1935), 256-57.
4. Deputy, E. C., "Knowledge of Success as Motivating Influence in College Work," J. Ed. Research, 20 (1929), 327-34.
5. Jones, J. E., "Experimental Studies of College Teaching," Archives of Psychology, No. 68, Sections IV and V (1923), 36-54.
6. Keys, N., "The Influence on Learning and Retention of Weekly as Opposed to Monthly Tests," J. Ed. Psychol., 25 (1934), 427-36.
7. Pease, G. R., "Should Teachers Give Warning of Tests and Examinations?" J. Ed. Psychol., 21 (1930), 273-77.
8. Sproule, C. E., "Suggestive Effects of the True-False Tests," J. Ed. Psychol., 25 (1934), 281-85.
9. Turney, A. H., "The Effect of Frequent Short Objective Tests upon the Achievement of College Students in Educational Psychology," Sch. and Soc., 33 (1931), 760-62.
10. Weissman, S. A., "The Effect of Frequent Testing on Achievement in High School Physics," Master's 1934, College of the City of New York, 41 p. ms.

INDEX

Absence, 7, 10, 28
Deputy, E. C., quoted, 1
Examinations, Research on
  Correction
    Sproule, C. E., 14
  Effectiveness
    Jones, J. E., 5
  Frequency
    Deputy, E. C., 1
    Keys, Noel, 3
    Smeltzer, C. H., 4
    Turney, A. H., 3
    Weissmann, S. A., 4
  Surprise
    Corey, S. M., 5
    Pease, G. R., 2
Findings, 18
Groups, Equating of, 7
Identical Factors, 14
Intelligence Factor, 25
Precautions, 15, 16
Problem and its Setting, 5, 6
Problems Suggested, 26, 27
Public School Publishing Company, 7
Recency, Check on, 22
Review of Research, 1
Reviews, 15
Stephens, J. M., 28
Study, Points of, 16, 17
Subjects, Selection of, 7
Technique, 13
Testing, 6, 7
Tests, Student Correction of, 14, 15
Time of Experimentation, 7
World Book Company, 7, 8

VITA

Sister Felicita Gable, Sister of Charity of Saint Vincent de Paul, was born in Mount Carmel, Pennsylvania, April 2, 1899. She received the degree of Bachelor of Arts, Saint Joseph's College, Emmitsburg, Maryland, June 1919. Her teaching commenced in 1928, when she substituted in Saint Mary's Academy, Norfolk, Virginia. The following year she taught Latin and English in the Saint Joseph's High School, Emmitsburg, Maryland. From September, 1929, to June, 1934, she taught science and mathematics in the Saint Joseph's High School, Portsmouth, Virginia. Throughout this teaching period, the writer pursued summer courses in higher education. In October, 1934, she entered the School of Higher Studies in Education of the Johns Hopkins University as a candidate for the Degree of Doctor of Education.